

# Applied Behavior Analysis Is Ideal for the Development of a Land Mine Detection Technology Using Animals

B. M. Jones

Behavioral Technology Group at the Shriver Center  
University of Massachusetts Medical School

The detection and subsequent removal of land mines and unexploded ordnance (UXO) from many developing countries are slow, expensive, and dangerous tasks, but have the potential to improve the well-being of millions of people. Consequently, those involved with humanitarian mine and UXO clearance are actively searching for new and more efficient detection technologies. Remote explosive scent tracing (REST) using trained dogs has the potential to be one such technology. However, details regarding how best to train, test, and deploy dogs in this role have never been made publicly available. This article describes how the key characteristics of applied behavior analysis, as described by Baer, Wolf and Risley (1968, 1987), served as important objectives for the research and development of the behavioral technology component of REST while the author worked in humanitarian demining.

*Key words:* applied behavior analysis, research and development, behavioral technology, land mines, detection, dogs

Domestic dogs (*Canis familiaris*) have long been using their keen sense of smell to locate a wide variety of target items for their human handlers (see Lorenzo et al., 2003). Some of the more unusual targets include invasive plant species (Goodwin, Engel, & Weaver, 2010), dairy cows in estrus (Kiddy, Mitchell, Bolt, & Hawk, 1978), endangered animal species (Cablak & Heaton, 2006), melanomas (H. Williams & Pembroke, 1989), off-flavors in cultured catfish (Shelby, Schrader, Tucker, Klesius, & Myers, 2004), gas-pipeline leaks (Quaife, Moynihan, & Larson, 1992), dangerous molds in buildings (Kauhanen, Harri, Nevalainen, & Nevalainen, 2002), and people who had handled items found at crime scenes (Schoon, 1998). Arguably the most valuable targets, however, have been land mines and unexploded ordnance (UXO) left buried in the

ground following armed conflicts (McLean, 2003).

Using dogs for mine and UXO detection became popular when humanitarian mine action boomed after the end of the Cold War and the Anti-Personnel Mine Ban Convention (also known as the Ottawa Treaty) was signed in 1997. The reported success of mine-detection dogs (MDDs) by field operators, along with interest in new techniques for detecting explosives after the terrorist attacks of September 11, 2001, also spawned a literature on odor detection by dogs. Given that dogs have no innate interest in mines or UXO, and so must be trained to search for, and then indicate, their presence, the paucity of behavior-analytic research in this literature is surprising. Only a small group of behavior analysts at Auburn University's Canine Detection Research Institute (e.g., Waggoner et al., 1998; M. Williams & Johnston, 2002) have published research investigating methods of training MDDs. Instead, MDD training, testing, and deployment protocols vary greatly across organizations, and those protocols are based largely on traditions within the organization rather than

---

I thank Rick Fleming for his assistance and three anonymous reviewers for their helpful feedback on an earlier draft of this paper.

Address correspondence to Max Jones, Behavioral Technology Group, University of Massachusetts Medical School, Shriver Center, Waltham, Massachusetts (e-mail: brent.jones@umassmed.edu).

empirical data (see Fjellanger, 2003). Whatever the reason for the absence of behavior-analytic research in this area, this article makes a case for behavior analysis being extremely useful; in particular, for developing the behavioral technology required in a modified use of MDDs. The case I make is based on personal experience working in humanitarian mine action and draws on principles with which readers of *The Behavior Analyst* will be familiar. Before making that case though, some background regarding the arena in which I worked seems necessary.

Between 2006 and 2008, I worked for the Geneva International Centre for Humanitarian Demining (GICHD). GICHD is a nonprofit organization that is affiliated with the United Nations (UN). Its mission is to conduct research, disseminate technical information, develop operational standards, and provide expert evaluations of ongoing activities with a view to increasing the safety, effectiveness, and efficiency of mine and UXO clearance operations. And if ever an industry required improved techniques for getting the job done, it is humanitarian demining because the scale of the problem in about 70 countries is immense and far exceeds the scale of resources that are invested to solve it. Explosive remnants of war are perhaps the epitome of man-made pollution; not only do they kill and maim innocent people, but they also deny millions of people access to the land and natural resources that would undoubtedly improve their quality of life.

The current process of declaring land to be free of mines and UXO and so safe for use by local communities is painstakingly slow and extremely expensive. First, government authorities attempt to identify suspect hazardous areas (SHAs) by consulting records of where armed conflicts and mine- and UXO-related accidents have occurred. Those authorities then either deploy their own

operators to the field or contract with other operational organizations (non-governmental organizations or commercial companies) to complete the task. The operators begin by sending out small teams to conduct general surveys in the various SHAs. These teams seek detailed information about the SHAs from local inhabitants with a view to defining smaller SHAs and marking those areas. Depending on the size and geography of the new SHAs, operators then send to the field either clearance assets (e.g., tillers, rollers, flailing machines) or further detection devices for technical surveys. Various devices are used for technical surveys, including an array of metal detectors and ground-penetrating radars carried by armored vehicles, MDDs, and mine-detection rats (see Poling, Cox, Weetjens, Beyene, & Sully, 2010). However, the most common technique involves recruiting local people and training them in manual demining. Manual deminers work alone in lanes marked out in the SHA (usually 1 m by 25 m). They are fitted with personal protective equipment, and use a metal detector, a prod, and various handheld excavation tools. On each occasion that their metal-detector sounds an alarm, they are required to step backwards some minimum distance, kneel on the ground, and use their prod and excavation tools to find the object that was sensed by the detector. This often involves very carefully excavating a hole that is about 1 m<sup>2</sup> and 50 cm deep, and so removing around 0.5 m<sup>3</sup> of sand or soil to a position behind them. Lanes that contain many metallic fragments or sand or soil that is rich in iron content can, therefore, take many hours (sometimes days) to inspect. Careful inspection is, however, paramount because many antipersonnel mines have minimal metal content and so are difficult to detect. After an SHA has been deemed to be free of mines and UXO, some system of quality control

involving other detection technologies is applied before the SHA is canceled and reported as such to the sponsoring authority.

#### **DEVELOPING MORE EFFICIENT DETECTION METHODS: REMOTE EXPLOSIVE SCENT TRACING**

Given the slow speed, high cost, and danger associated with existing clearance methods, GICHD (and the humanitarian demining industry generally) have long been interested in developing more efficient methods for detecting mines and UXO. One such candidate, and that which I was asked to investigate, is known as remote explosive scent tracing (REST). REST refers to a method for detecting areas of land that contain chemical evidence of mines or UXO; in particular, areas that could be defined as *minefields*. In its generic form, REST involves collecting samples of air or dust from defined areas within SHAs (a sampling phase) and presenting those samples to either mechanical or animate detectors in a remote location (an analysis phase). The areas corresponding to samples that are judged to be positive by the detectors are then either searched more thoroughly by technical survey methods or cleared by machines. In contrast, the areas corresponding to samples that were judged to be negative by the detectors are generally exempt from further inspection. Thus, an accurate and reliable REST system would constitute a quick and relatively inexpensive strategy for distinguishing hazardous from nonhazardous areas, and so facilitate the spending of scarce resources in only those areas where they were actually needed. (Those who work in humanitarian demining often say that 99% of the area they work in turns out to be free of mines; if only they could have identified the relevant 1% of land before investing their resources.)

In addition to canceling nonhazardous areas rapidly, a REST system appears feasible for two main reasons. First, a company based in South Africa (Mechem Ltd.) reported that they had already developed an operational REST system using dogs (Joynt, 2003) and they were winning UN contracts with it. Second, the animals in a REST system are required to perform a signal-detection task that seems fundamentally similar to the task facing dogs and rats that are used in known minefields to pinpoint mines or UXO (i.e., direct detection animals), and numerous organizations have reported considerable success with these animals (McLean, 2003; Poling, Cox, et al., 2010). The differences between remote and direct detection methods appear to lie only in where the inspection of samples takes place (i.e., a laboratory vs. the field) and in the strength of the signal presented to the animal. (Animals that work in REST are likely to be presented with positive samples that contain considerably less chemical residue, and thus less odor, of mines and UXO than are direct detection animals, because the contaminated air or dust from immediately above a mine or UXO will have been diluted by the uncontaminated air or dust collected from areas between mines or UXO.)

In early 2006, GICHD, together with Norwegian Peoples' Aid (NPA) and Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling (APOPO), constructed laboratory facilities in Morogoro, Tanzania for the purpose of developing REST. (This facility was located adjacent to APOPO's facility on the grounds of the Sokoine University of Agriculture.) NPA relocated 10 dogs and various apparatus to Morogoro from their Angola operation; 17 locals were employed to staff the facility, and I recruited an expatriate research assistant with a masters' degree in behavior analysis.

Because all 10 dogs had prior training in an earlier attempt to

develop REST, our first set of procedures was a stripped-down version of the set that the dogs had most recently experienced. Briefly, six dogs received two training sessions per day. In each session, a dog was taken in and out of a room that contained an apparatus called a *carousel*—a 12-spoked wheel suspended horizontally about 50 cm above the floor and with cups at the end of each of spoke. Each cup presented a sample to the dog. Positive samples consisted of a measured amount of 2-4-6 trinitrotoluene (TNT; the explosive in most UXO and mines) dissolved in water and added to 10 g of sand in an aluminum pot similar to a film canister. Negative samples consisted of an equal amount of water added to 10 g of sand in these pots. Each dog received eight sets of 12 samples per session in which 10 were positive and 86 were negative. The indication response was sitting. Thus, in signal-detection terminology (Green & Swets, 1966), sitting immediately after sniffing a positive sample was a *hit*, walking to the next sample after sniffing a negative was a *correct rejection*, sitting after sniffing a negative was a *false alarm*, and walking past a positive was a *miss*. Neither correct rejections nor misses earned programmed consequences. However, a hit was followed by a clicking sound, the delivery of a food treat, and removal from the room, whereas false alarms earned removal and a brief period of time-out in some conditions. A handler released the dog to search, stayed inside the room while the dog was searching, delivered food for hits, and led the dog from the room. The handler received instructions from a laboratory technician viewing the trial from outside the room through a one-way mirror, and a scorer recorded for each sample the technician's decision as to whether a dog had sniffed it and whether it had sat after sniffing. All samples that were judged to have been sniffed on a visit to the room

were removed before the next visit, and a new set of 12 samples was presented after all samples in a previous set had been sniffed.

### THE NEED FOR RESEARCH THAT USES THE PRINCIPLES OF ABA

Prior to my employment with GICHHD, experts with experience training dogs for odor-detection roles had been contracted to develop the training, testing, and operational procedures for animal-based REST. (An ecologist and a psychologist were, however, employed to assist with testing their systems.) These experts began with Mechem's protocols, modified them to fit their opinions of the best methods, but reported only limited success when their systems were tested (see Fjellanger, Andersen, & McLean, 2002). Consequently, my first task at GICHHD was to convince my supervisor that empirical research rather than expert opinion was necessary for the development of REST. (It was interesting that he appeared surprised to hear me confess that I did not already know how to set up the system.) My argument went as follows. If an effective system of REST was to exist, it would share at least two features with any system engineered to detect environmental compounds more accurately than an unaided human: (a) Both systems would undoubtedly consist of sets of complex procedures, and (b) those procedures would have been discovered by rigorous and replicable empirical research. In the case of an electronic detection device, few people would challenge the need for sophisticated research and development (R&D) by scientists and engineers. However, fewer people recognize a similar need for thorough R&D in the preparation of animals for REST. Perhaps this is because most people involved with odor detection by animals are unaware

that there exists a science of behavior that is presently being used to generate training and behavioral control technologies in other domains. Whatever the reason, I argued that the complexity of procedures in an operationally viable system of REST would likely be equal to (if not greater than) the complexity of procedures that culminate in an electronic device. It therefore follows that the need for scientific research in the development and validation of procedures in REST should be given equal recognition.

After I had obtained my supervisor's support for taking a more research-focused and slower approach to developing REST, I then set about coaching him and other colleagues on the basic features of applied behavior analysis (ABA). Much of that information is very familiar to readers of this journal and I will not re-present it here. Suffice to say, I gave them Cooper, Heron, and Heward's (1987) definition of ABA, described the various professional bodies in our field, and mentioned some of the domains in which ABA has been used. Some of the people with whom I worked neither needed nor wanted any more information about the approach to REST R&D that I would be taking, and I accepted that. There was, however, a group that I believed should learn more, because the quality of our R&D depended on their understanding some key principles of ABA. This group included scientists and dog-training professionals whom I recruited to serve on an advisory committee and some of the local staff we employed as technicians. I routinely organized those principles in terms of Baer, Wolf, and Risley's (1968, 1987) seminal articles.

Baer et al. (1968, 1987) described seven key characteristics of ABA; it is applied, behavioral, analytic, technological, conceptually systematic, effective, and addresses generality. These characteristics can also be

considered criteria for behavioral research that seeks empirically validated procedures for establishing specific performances in animals and people (i.e., research that seeks to develop a behavioral technology). More important, I proposed that the R&D of procedures for preparing dogs as detectors in a REST system would be most productive if the Baer et al. characteristics were used as objectives. I therefore described the Baer et al. characteristics of ABA in varying degrees of detail. The full and formal version of that description is below and should serve to illustrate just how well suited ABA is to the development of REST.

*ABA and REST R&D both involve applied research.* Baer et al. (1968, 1987) argued that behavior-analytic research is *applied* if the target behavior of interest is socially significant and is, therefore, either directly beneficial for the animal or person emitting the behavior or valued by the community in which the animal or person exists. This definition of *applied* was intended to distinguish ABA from the experimental analysis of behavior (EAB) where basic research aimed at investigating principles of learning and behavior involves studying a target behavior that has been selected for measurement convenience (e.g., key pecking in pigeons). Baer et al. (1968) wrote that "the differences between applied and basic research are not differences between that which "discovers" and that which merely "applies" what is already known" (p. 91). Instead, the purpose of both applied and basic research is the same; namely, to discover lawful relations between an animal's experience with environmental events (its learning history included) and its current behavior. It is only the immediate social significance of the behavior being studied that determines whether it is deemed applied or basic research. Therefore, to the extent that a dog sitting reliably in the presence of mine or

UXO odor is of value to people seeking to remove that mine or UXO from the environment, that behavior is socially significant, and research aimed at establishing and studying it is *applied*. This definition of applied research and the associated concept of a basic–applied continuum of research renders a distinction between *research* and *development* (two arch enemies to many) problematic because the two activities are seen as one and the same; the development of a behavioral technology for animals serving in a REST system necessarily involves research activity.

*ABA is, and REST R&D should be, behavioral.* Baer et al. (1968, 1987) stated that ABA is *behavioral* to the extent that it involves precise and objective measurement of observable aspects of an individual's behavior and the environmental conditions under which a specific response occurs. Various strategies for recording that a response has been emitted are described in the behavior-analytic literature (e.g., some responses permanently change the environment, thus permitting permanent-product recording; some responses can be sensed electronically), but observation of the subject behaving through time is often the only method available when the target behavior is socially significant. Indeed, in REST systems, the indication response (sitting in dogs or scratching in rats) has always been detected and recorded by human observers. (Interestingly, REST with animals should, therefore, be viewed as two interacting signal-detection tasks, one being performed by the animal and the other being performed by human observers, and with considerable potential for topographical drift of the animal's indication response because the human's decisions define the contingencies of reinforcement received by the animal.) This involvement of another agent's behavior in the research implies a need to verify

objective recording of the animal's behavior, and that it was the animal's behavior, and not the observer's behavior, that changed with the introduction of some change in the training procedure. This verification generally proceeds by quantitative assessments of the degree of agreement between behavior recordings of two or more independent observers. However, another technique, and one that is suited to REST R&D, is to measure the agreement between a single observer's recordings before and after the observer is given information regarding the placement of samples containing the target odor (termed *blind testing*). Only when we have obtained high degrees of inter-observer (or intraobserver) agreement can we be confident that we have reliable data. As it happens, a significant investment in the documentation of procedures, the training of staff, and the retraining of dogs was required before we succeeded in obtaining high measures of interobserver and intraobserver agreement.

*ABA is, and REST R&D should be, analytic.* This characteristic of ABA is perhaps the one which has most often been absent in previous attempts to develop REST using animals, and its absence is possibly the reason why those attempts failed. There are at least three ways in which R&D should strive to be analytic. The first follows directly from Baer et al.'s (1968) definition of *analytic*. That is, the research should seek to demonstrate convincingly, and by way of controlled experiments, that any change in measures of the subject's target behavior is due to a specific and accurately identified change in the training procedure and not some other confounding variable. The second sense in which R&D should be analytic involves seeking objective and empirical evidence for the behavior we are trying to establish. The third meaning of analytic lies in paying close attention to how the target behavior is record-

ed so that analyses of the data might suggest controlling variables that would not otherwise be detected.

Requiring empirical evidence that a specific procedure produced the desired behavior change is important, because only then can we develop an understanding of the variables that control the target behavior and expect to develop an effective and replicable behavioral technology. In the context of REST R&D, a behavior analyst will assume that the animals (rather than any human expert) know best which procedures assist their detection accuracy and set about asking them a series of simple and clear questions. Consequently, the process of developing a set of training, testing, and operational procedures will be iterative and involve many procedural changes and comparisons. Although REST researchers ought to work toward procedures that are feasible for operational activity (e.g., multiple dogs will need to examine the same set of samples), progression toward those final procedures should, therefore, proceed systematically and by analyses of the effect, if any, of numerous procedural changes where the value of a single variable is changed in isolation. The importance of implementing, and then assessing the effects of, single and small changes to a procedure was seldom recognized in previous attempts to develop REST. Instead, numerous procedural changes were often made simultaneously, with a view to fixing all the perceived problems in one hit.

In addition to the value of seeking scientific evidence for the efficacy of training procedures, the actual research methods used in ABA to provide this evidence are ideally suited to preparing a small group of animals for operational REST activity. The research methodology used in ABA is called a *single-subject* (or *within-subject*) *experimental design*. It was pioneered in the behavioral sciences by Skinner (1938), elaborat-

ed by Sidman (1960), and continues to evolve (e.g., Kazdin, 2010). The basic principles of this methodology are as follows. Some specific behavior of each of a small number of subjects is first measured repeatedly in successive sessions that involve a specific and unchanging method of training, a period known as *baseline*. Measures of the behavior of each subject are then plotted with session number on the abscissa so that the stability of the target behavior over sessions can be assessed. Once the baseline measures are deemed to be stable and with acceptable variance, each of the subjects receives the same procedural change and the monitoring of their behavior continues. An effect of the procedural change is claimed if measures of the target behavior after the procedural change fall in an area of the graph that is outside the area in which they would be predicted to fall if the initial training conditions had remained. (I frequently used Cooper et al.'s, 1987, section on *baseline logic* to illustrate this method of inductive reasoning.) Thus, in contrast to between-subjects designs, in single-subject research, each subject receives the control condition (the baseline phase) and the experimental condition (the intervention phase). Consequently, each subject constitutes an experiment in its own right, and the value of having more than one subject is that it affords an opportunity to replicate the result of the experiment multiple times. Although it is possible that all the subjects share some preexperimental experience or some biological trait that modulates an effect of the procedural change, that possibility become less likely with each subject added. Five subjects (i.e., four replications of an effect) is generally considered sufficient to prove an intervention's efficacy. It was interesting that most of my colleagues were surprised to learn that meaningful research could be conducted with such a small number of animals; they were assuming that

all research had to involve group-design experiments and as many animals as possible per group.

The design we used most often was an A-B design (in which A refers to a baseline phase and B refers to the change in procedure) because many of our procedural changes either clearly improved the detection accuracies of all our dogs or had no effect but were nevertheless desirable changes from our standpoint. However, on several occasions we sought evidence that the mere passage of time was not responsible for the behavior change by staggering in time across subjects the point at which the procedural change was made, a design known as multiple baseline across subjects (Baer et al., 1968). We also occasionally used A-B-A withdrawal designs (Ruszh & Kazdin, 1981) to increase our confidence in experimental control over the target behavior and an accurate description of the variables that were responsible for the behavior change. (These designs were particularly useful for assessing intraobserver agreement via blind tests.) Finally, one experiment involved a parametric design in which quantitative measures of an independent variable were varied. Whichever design was used, posting in a public place session-by-session measures of the dogs' performances as they came to hand served to illustrate the design logic and generate greater interest in the research. Thus, all colleagues and staff witnessed the results of experiments (a) aimed at reducing false-alarm rates; (b) identifying the topography of stimulus control (see below); (c) investigating effects of varying the ratio of positive to negative samples; (d) assessing the effect of procedural changes that moved the system closer to one that was feasible for operational activity; and (e) assessing inter- and intraobserver agreement.

I often emphasized three points regarding single-subject designs. First, the various principles applied

when proving an effect of some procedural change can be used in a wide variety of ways, meaning that the methodology is conducive to improvised experiments that are designed on the fly as data are generated. Second, successive experiments using these designs can be conducted while moving the system closer to the set of procedures that will be employed in operational activity. We need only to preserve in the training protocol those procedural changes that improved our animals' detection accuracy (or had no effect) and repeatedly redefine the set of procedures being considered baseline. Third, to easily compare data sets across conditions (sets of sessions) in a single animal and judge accurately when a real change in behavior has occurred, the variability in measures of behavior across sessions within a condition must be minimized. This minimization of variability is achieved by minimizing observers' measurement error and keeping at a constant value across sessions all those variables that have some effect on the probability of the occurrence of the target behavior. However, controlling important variables itself requires knowing which variables affect the emission of the target behavior and so requires that efforts to discover those variables are part of the research agenda (see below).

There is a second but equally important aspect to being analytic in the R&D of training procedures for REST animals. It concerns seeking empirical evidence that the target behavior we are trying to establish in our animals (i.e., sitting in the presence of UXO or mine odor) is indeed the behavior that we have trained. (A discussion of stimulus control topography coherence theory [SCTCT] is relevant here; see McIlvane & Dube, 1992, 2003). Our goal is to have this response under stimulus control by only UXO or mine odor, such that the response occurs on every occasion that a sniffed sample contains UXO

or mine odor (i.e., a 100% hit rate), but never when a sniffed sample does not contain that odor (i.e., a zero false-alarm rate). This is not a trivial exercise, because other irrelevant odors can easily occur with the intended target odor and come to cue the indication response (a phenomenon known as *overshadowing*) without our noticing that this has happened. In terms of SCTCT, the topography of the stimulus control that exists over the indication response might not be the topography that we intended.

The problem of inappropriate stimulus control arises because we are trying to establish this control by a stimulus that we, as human trainers, are unable to sense, and this inability in turn reduces our ability to arrange differential reinforcement of the response with respect to the presence and absence of the target stimulus. (Note that we were often training with amounts of TNT that were lower than the detection thresholds of analytic chemistry instruments such as an ion scan.) The problem is not unlike the problem that a parent faces when trying to teach his or her child to label some emotion like pain, fear, anxiety, envy, and so on; the appropriate stimulus cannot be sensed by the parent (it is a private event for the child) so that the parent can reinforce an appropriate verbal report by the child. Instead, the parent must be guided by collateral behavior (e.g., the child complaining of a toothache holds her jaw) or environmental antecedents to the behavior (e.g., the child reporting fear having just experienced a dog growling at him). Similarly, the behavior analyst who attempts to train stimulus control by an odor that is beyond verification should seek empirical evidence for the intended stimulus control in results other than just high hit rates and low false-alarm rates in training. We cannot rely on our sample-manufacturing methods resulting in the target odor being

present on all positive samples and absent on all negative samples, and the only odor that was predictive of reinforcement for the indication response.

Unfortunately, seeking evidence for stimulus control by a specific target odor is a complex matter and requires creative experimentation. We took two approaches. First, in numerous blocks of sessions, we used a titration procedure to manufacture five concentrations of TNT in the 10 positive samples per session and reasoned that the odor of TNT (or its breakdown products) on positive samples likely had stimulus control if hit rates fell with decreasing concentration. Second, we conducted numerous experiments aimed at disproving stimulus control by odors other than TNT. Those experiments generated mixed results. For example, after only a week of training our dogs to indicate the presence of TNT, the results of a test revealed that the dogs were in fact indicating the presence of acetone on positive samples because acetone had been used in the cleaning process after preparation of positives but not negatives. Similarly, a later test revealed that odors acquired from the location where positive samples (but not negative samples) were made had acquired stimulus control in all our dogs. Although these results were setbacks in the pursuit of an operational system, they would never have been obtained had we been satisfied with high detection accuracies in training and not maintained skepticism about the topography of the stimulus control we had trained.

Finally, productive R&D in REST should be analytic in the sense of conducting thorough analyses of measures of the indication response with a view to discovering variables that control the emission of that response. These data analyses are possible only when the indication response is recorded with regard to many potentially important vari-

ables. For example, rather than simply counting the number of hits and false alarms per session, indications should be recorded as a function of the spatial position of samples in a set, the temporal position of samples in a session, the observer who judged the animal's behavior, the technician who created the samples, climatic conditions inside the laboratory where samples are presented, the number of times those samples had been presented previously, and so on. It is only after the values of many variables have been recorded alongside records of the animal's behavior that one can search for correlations between values of a variable and measures of the animal's behavior. Taking such an approach, we found clear evidence (a) of a *warm-up effect* (Sidman, 1958) where hit rates were lower on the first two sets of samples in a session than they were on the remaining six sets; (b) that dogs showed higher hit rates in the second session of a day than in the first; and (c) that the hit rate of most dogs was highest when they were the first to inspect a set of samples. We were able to identify the behavioral processes underlying some of these phenomena. However, they also constituted sources of variance in measures of the target behavior and could, therefore, guide strategies aimed at either accommodating or minimizing that variance.

The warm-up effect serves as a useful example of a source of variability in session-by-session hit rates of an animal. We used a computer program to randomly allocate 10 positive and 86 negative samples to Positions 1 through 96 in a session. If hit rates on the first two sets of samples (Positions 1 through 24) were always lower than the remaining sets, and the program happened by chance to allocate many of the 10 positive samples to the first two sets, then the overall hit rate for the session would be lower than it was in a session in which relatively few

positive samples appeared in the first two sets. Although it would be unwise to reduce the number of positive samples early in a session to reduce hit-rate variability, technicians who arrange operational samples would be wise to re-present later in a session those samples that had been presented early in a session.

*ABA is, and REST R&D should be, technological.* That ABA is technological means that all important aspects of a specific procedure for changing behavior (including procedures for establishing the appropriate stimulus control over some response) have been completely and accurately identified and described (Baer et al., 1968). When ABA research is published, complete and accurate descriptions of successful techniques for generating desired performances in a subject allow others to apply those techniques and achieve similar outcomes (i.e., being technological is a prerequisite for a behavioral technology). Thus, teaching children with autism to speak or training animals to indicate the odor of mines is not considered to be an art, or a skill possessed only by experts (cf. McLean, 2003). Instead, it is considered to be a science and, as such, requires careful application of techniques that others have documented thoroughly and found to be effective. Furthermore, being technological should be considered to be an objective of research rather than a dichotomous feature, because the completeness and accuracy of procedural descriptions often evolve in the course of experimentation. In other words, much of what is learned in the assessment of behavior-change techniques is exactly what aspects of a procedure are critical, what aspects are important, and what aspects are superfluous.

The need for R&D into REST with animals to be technological was implied in the mission I was given when joining GICHD—to discover, document, and disseminate how to

prepare animals for REST. However, the identification and description of those procedures are necessary not only for other people interested in replicating the success but also for technicians who conduct the research, because they need to replicate procedures across training sessions in a given condition to minimize the variability in measures of a dog's performance. Unfortunately, accurate description of procedures for research assistants was a formidable task, because those procedures were sometimes extremely complex. For example, there were five types of research assistants working on our project: laboratory technicians, scorers, dog handlers, cleaners, and security guards. Furthermore, each type of assistant had numerous sets of tasks (e.g., the laboratory technicians collected materials to be used as odorants, manufactured samples for training sessions on the next day, sorted and stored samples, presented and removed samples in a training session, judged the dog's behavior, etc.). To promote procedural integrity across assistants and across time in a single assistant, we wrote standard operating procedures (SOPs) for each type of assistant, and diligently updated those SOPs prior to every procedural change that constituted a new condition in an experiment. Those written only for the laboratory technicians exceeded 45 pages and included such details as the total number of times that a sample should have been touched prior to presentation, and exactly when in a training session a technician should change his or her outer pair of plastic gloves. We also found documentation to be insufficient. Instead, it proved necessary to run regular SOP training sessions for each group of technicians, examine their knowledge of current SOPs in short tests, develop checklists for supervising technicians to assess the performance accuracy of others, and devise systems for supervising the supervisors. Variability in

the performance of even our security guards had the potential to add variability to our dogs' indication accuracies and so hinder the progress of our research, because we were using food rewards in training sessions and those guards were required to give our dogs a measured amount of food at 5:30 a.m. each day. Thus, if time since the last meal (or size of last meal) determines the efficacy of food rewards in training sessions, then adding variability to meal times (or meal sizes) will add variability to reward efficacy and perhaps also, as a consequence, add variability to measures of the dog's performance.

*ABA is, and REST R&D should be, conceptually systematic.* A conceptually systematic body of research is one in which methods are described and results are discussed in terms drawn from a conceptual framework that has evolved with the maturation of that body of research (Baer et al., 1968). For the most part, published examples of ABA have been conceptually systematic insofar as the procedures for establishing a specific behavior and the mechanisms that underlie the efficacy of those procedures have been described in terms of the basic principles of learning and behavior involved. In contrast, the literature describing methods for training odor detection in dogs is far from being conceptually systematic (McLean, 2003). Furthermore, Baer et al. argued that being conceptually systematic "can have the effect of making a body of technology into a discipline rather than a collection of tricks" (p. 96), and various published descriptions of odor-detection training methods resemble "a collection of tricks" largely because there is little consensus over an appropriate vocabulary of technical terms. (For example, Cablk & Heaton, 2006, saw it necessary to define terms such as *body language*, *prey drive*, and *hunt drive* when describing their methods.) Although research concerned with odor-detecting dogs is far less mature

than ABA, I argue that that research should strive to be more conceptually systematic and that ABA offers an appropriate conceptual system.

The conceptual framework adopted in ABA has its roots in two domains: (a) basic research that investigates the principles of learning and the maintenance of learned behavior in animals (EAB), and (b) a philosophy of psychology known as radical behaviorism (Skinner, 1953). Only a brief description of both disciplines should suffice to support my argument that the conceptual framework adopted by researchers who conducted REST R&D should also be guided by EAB and radical behaviorism.

Radical behaviorism assumes that all aspects of an animal's behavior are determined; that behavior is lawful. This philosophy challenges the scientific value of so-called mentalistic explanations of behavior; that is, explanations that refer to thoughts, emotions, states and drives inside the animal. Instead, three sets of factors are assumed to determine the behavior of an animal at any given time: (a) its biological make-up, including what it has inherited from its ancestors, (b) its learning history, and (c) the current state of the environment. For example, rather than attributing a dog's false alarm in REST to its *thinking* that it smelled explosives or to too high a prey drive, a radical behaviorist might attribute the false alarm to the absence of reinforcement for correct rejections, and of course then conduct experiments to test that hypothesis. A further assertion is that the utility of some proposed mechanism for how Variable X affects measures of Behavior Y should be judged in terms of the degree to which that mechanism allows one to predict and control measures of Behavior Y. This is a philosophy of science known as *pragmatic epistemology*, and its adoption suits any field of research that seeks real-world applications of principles discovered.

EAB is directly relevant to REST R&D because EAB researchers often study learning in nonhumans, and the behavior analyst who works on REST must discover how best to train (i.e., facilitate learning in) his or her animals. Although several authors have recently brought EAB to the attention of some animal trainers (most notably, Pryor, 2002), only elementary principles have been described (e.g., Fjellanger, 2003; Hilliard, 2003), and few animal trainers appreciate just how pertinent the EAB literature actually is. For example, of those aware of EAB, only a small proportion will be aware of the research that has been concerned with variables that control when, and at what speed, animals learn to discriminate (i.e., respond differentially to) environmental stimuli. Among other questions, this research has examined the factors that predict which feature of a complex multi-featured stimulus will earn an animal's attention and so acquire stimulus control over some response (see Johnson, 1970, for an early review). Termed *selective stimulus control*, various phenomena studied in this literature are relevant to REST R&D (e.g., blocking, overshadowing, masking) because the odor of mines and UXOs is probably a bouquet with many constituent odors, and being able to control exactly which odor feature controls the animal's indication response would be of great benefit when attempting to establish the appropriate performance of our animals. Granted, much of this research in EAB has focused on visual discrimination in pigeons (e.g., Farthing & Hearst, 1970; Miles & Jenkins, 1973), but research conducted with mammals and olfactory stimuli has also appeared in the literature (e.g., Laing, Panhuber, & Slotnick, 1989; Slotnick & Katz, 1974), and that research has replicated many of the findings with pigeons. Thus, numerous principles of selective stimulus control, and of learning generally,

seem to apply across a range of animal species and likely also apply to dogs being trained for REST.

Furthermore, although only a small number of animal trainers involved with REST are familiar with the EAB literature, an even smaller group of trainers is aware that there is a subdiscipline within EAB that is concerned with developing methods to measure the sensory abilities of various animal species (a field known as *animal psychophysics*) let alone being aware of the principles discovered in that field (see Stebbins, 1970, for an early review of this field). Out of necessity, researchers who seek to quantify the sensory thresholds of animals must be informed by findings from animal studies of discrimination learning (and variables that control asymptotic performance) because they need to establish and then maintain behavior that is maximally sensitive to the strength of a specific stimulus. Similarly, researchers who attempt to develop animal-based REST need to be informed by animal psychophysics research because accurate determination of an animal's threshold for sensing an odor (or odors) indicative of mines and UXO should be one of the most important goals of their R&D. That is, the concepts validated and the technical jargon used in EAB studies of selective stimulus control and animal psychophysics would ideally feature in the conceptual framework adopted by scientists who conduct REST R&D. Admittedly, contemporary research in these areas is extremely sophisticated and highly technical, and requires considerable formal education. Therefore, at the very least, scientists with expertise in animal psychophysics should serve on a scientific advisory committee that oversees REST R&D.

*ABA is, and REST R&D should be, effective.* Baer et al. (1968) wrote that "if the application of behavioral techniques does not produce large enough effects for practical value,

then application has failed" (p. 96). Furthermore, they argued that "in evaluating whether a given application has produced enough of a behavioral change to deserve the label [effective], a pertinent question can be, how much did that behavior need to be changed?" (p. 96). Obviously, this question is meant to be a purely practical one, and as such, its answer is expected to be supplied by consumers of the research. In the case of animal-based REST, R&D efforts will be judged to have been effective if the minimum performance requirements specified in the International Mine Action Standards (IMAS: 09.43; Remote Explosives Scent Tracing) have been met when operational samples are analyzed. The 2005 revised version of those standards specifies that (a) each individual animal must achieve a minimum hit rate of 70% and a maximum false-alarm rate of 5%; (b) at least three animals must be used to inspect the same set of samples; and (c) as a group, they should achieve a cumulative hit rate of 100% and a cumulative false-alarm rate below 20%. In other words, the system will be judged to be effective when no positive samples have been missed by the group and no more than 20% of negative samples have been indicated.

Although IMAS 09.43 describes clear criteria for judging whether a REST system is effective, and being effective is an important criterion when judging the success of R&D efforts, the practical utility of a REST system should not be judged by accuracy criteria alone. Instead, the efficiency of the system must also be considered, and this requires comparing it to alternative technologies that exist for detecting mines and UXO. (In turn, the resources required for any method of detection must themselves be evaluated against those required for simply clearing areas using mechanical devices such as tillers and flails.) For example, we

know that leading appropriately trained dogs or rats to a known minefield will usually result in their locating the mines, at least down to the nearest square meter. It therefore follows that collecting the dust from a  $1\text{-m}^2$  area above the mine and presenting that dust to a trained dog in a remote laboratory will likely result in that dog indicating the presence of mine odor also. However, little, if any, efficiency will have been gained by the remote sensing system. Clearly, if REST is to be more efficient than any direct detection method, then the land area represented in a sample of dust must be considerably larger than  $1\text{ m}^2$ , but exactly how large that area needs to be is still yet to be specified by experts in humanitarian demining. (Of course, efficiency does not increase as a linear function of sample-area size, because more resources will be required to follow up on and reexamine larger areas that the REST system deemed to be contaminated with mines or UXO.)

In addition to requiring minimum sampling areas before being able to judge efficiency, it is only after these areas have been specified that the feasibility of a REST system can be properly assessed. The issue is a simple one but one that is worth illustrating. Suppose that there exists only one mine in a number of minefields that each measure  $1,024\text{ m}^2$  and we collect just one sample of dust from each field. (An area of  $1,024\text{ m}^2$  was chosen because it is between the  $230\text{ m}^2$  envisioned as practical by NPA, Fjellanger, 2003, and the  $8,000\text{ m}^2$  being sampled in operational theaters by Mechem Ltd., Joynt, 2003.) We always start by collecting dust from immediately above the mine but collect from increasingly larger areas across minefields. If our smallest collection area is  $1\text{ m}^2$ , then with each doubling of the area sampled (i.e.,  $2\text{ m}^2$ ,  $4\text{ m}^2$ ,  $8\text{ m}^2$ , etc.), the concentration of contaminated dust in our sample will

have been halved. Thus, in a sample taken from an area of  $1,024\text{ m}^2$  (i.e., the entire minefield), the concentration of contaminated dust will be  $0.5^{10}$  ( $0.0009 = 0.09\%$ ) of what it was in our sample from  $1\text{ m}^2$ . Whether or not the animals in a REST system can detect the presence of mine-related odors in that sample of  $1,024\text{ m}^2$  amounts to asking whether that sample contains a signal strength that exceeds the animal's sensory threshold, because no amount of optimized training can render that threshold irrelevant. Unfortunately, the research conducted at our facility never matured to a point at which this critical question could be asked of our dogs.

*ABA does, and REST R&D should, address the issue of generality.* In ABA, "a behavioral change may be said to have generality if it proves durable over time, if it appears in a wide variety of possible environments, or if it spreads to a wide variety of related behaviors" (Baer et al., 1968, p. 96). Similarly, the R&D of an animal-based REST system should strive to show generality in the animals' behavior change that has been induced by training. In most applications of ABA, the behavior change will involve a person now behaving in closer accordance with societal and cultural norms (e.g., no longer throwing tantrums or now asking to use the bathroom), whereas in REST R&D, the behavior change involves the animal now emitting the indication response only in the presence of the target odor. Despite this difference, in both ABA and REST R&D, a performance that has been established by training is required to be maintained across both time and situations in order to be useful. This maintenance, in turn, requires (among other things) that some degree of stimulus generalization occurs. That is, stimuli that are similar to those that were presented as the target stimuli in training but were never actually presented during

training should evoke the taught performance.

In animal-based REST, stimulus generalization circumvents the need to present in training all possible variants of mine and UXO odors at all possible strengths. Furthermore, attempting to minimize the requirement for stimulus generalization by training with target odors (i.e., using training aids) that are as close as possible to those that emanate from positive operational samples is rife with practical difficulties. For example, this approach requires that the R&D of the sampling phase of REST is sufficiently mature to offer precise descriptions of collection, storage, and presentation methods, but this may not be the case. In addition, positive samples would have to come from numerous actual minefields that were discovered by minimal use of detection technologies and so unlikely to contain extraneous odors that were added in the process of inspection. (Thus, one could not use soil collected from minefields that were discovered by MDD because those areas are likely to contain some amount of odor left by the dogs.) Similarly, negative samples would be required from many areas that were previously considered hazardous but turned out, again after minimal inspection and spoiling, to be entirely free of UXO. Finally, there is the problem of verifying the presence of mine or UXO odor in samples collected from the field when current instruments lack the sensitivity required.

After a researcher who conducts R&D in animal-based REST recognizes the requirement for stimulus generalization, he or she must then select target odors for training and address via experimentation how best to train and test for maximum generalization from those targets. Again, principles discovered in the EAB and in ABA imply various strategies, some of which have been applied by professionals training detection dogs, albeit in the absence

of our conceptual framework. The first strategy involves attempting to identify the specific odors that are common to a range of mines and UXO, and training stimulus control by those odors so that that control might later block the development of control by odors that temporarily accompany mine or UXO odor but are actually irrelevant. Thus, for a period of time, we tried to establish stimulus control by the odor of TNT or its breakdown products (e.g., 2-4 dinitrotoluene, DNT), because TNT is an ingredient in many types of mines and UXOs that differ in terms of other constituent compounds. Several researchers had assumed that this ingredient provided the odor signature of mines to dogs (i.e., Johnston et al., 1998; Phelan & Barnett, 2002), and Mechem Ltd. had been using TNT as their training aid. However, an attempt to apply blocking will undoubtedly result in the greatest generalization to operational samples if the choice of a target odor is based on the results of transfer tests (Reynolds, 1961) with animals that have proven their ability to detect mines and UXO in direct detection scenarios. Briefly, a transfer test involves assessment of the stimulus control exerted by each of a number of features of a compound stimulus by presenting each feature alone and measuring the degree to which the target behavior is emitted. Therefore, a transfer test to assess which component of a specific mine had stimulus control might involve burying the explosive content alone, the plastic casing alone, the metallic components alone, and the rubber seal alone, and then comparing the number of occasions that a dog (or several dogs) correctly indicates the position of each component. Such tests are difficult and have yet to be formally conducted, largely because sourcing the separated components of mines is problematic.

A second strategy for promoting stimulus generalization from a target stimulus and one that is known to most behavior analysts is to train with multiple exemplars of that stimulus (Stokes & Baer, 1970). In the context of REST, this involves training with positive samples that include odors of many different types of mines and UXO, those odors at different strengths, and those odors accompanied by a range of irrelevant odors. Similarly, negative samples should include a range of irrelevant odors to guard against an animal simply learning to reject a specific odor and indicate on all others (M. Williams & Johnston, 2002). Such training is likely to be effective because it essentially reduces the degree to which successful operational activity requires stimulus generalization. Again, however, sourcing those mines and UXO poses a significant barrier to this approach because transport of such items is highly regulated.

A third strategy is to consider the animals' task as one of learning to discriminate two large sets of stimuli (a set of odors associated with mines and UXO and a set of odors that are unrelated to mines and UXO) and, thus, learning the *concept* of a mine or UXO. Concept learning in animals has received considerable attention in EAB and is quite well understood. For example, pigeons have been able to learn to discriminate the presence versus absence of people in photographs (Herrnstein & Loveland, 1976), Bach's music versus Stravinsky's music (Porter & Neuringer, 1984), and impressionist art versus cubist or abstract art (Watanabe, Sakamoto, & Wakita, 1995). In terms of REST, this strategy involves giving the animal more freedom to identify the odor feature that distinguishes the set of positive samples from the set of negative samples, and so circumvents the need to identify which particular feature of mine or UXO odor ought to have stimulus

control (cf. the blocking strategy). However, the downside of this strategy is that by relinquishing control over the feature of positive samples used by the animals, extremely large sets of positive and negative samples will likely be necessary before stimulus control by the concept of mine or UXO (and not control by irrelevant odors temporarily correlated with them) develops. In addition, frequent assessment of the topography of stimulus control will be paramount, and a systematic approach to presenting stimulus sets will probably be necessary to limit the degree to which stimulus control by irrelevant odors overshadows control by mine and UXO odors.

Although some progress toward developing methods for training stimulus generalization was made at our facility (including a method of combining many irrelevant odors with TNT), we conducted only a limited amount of testing with samples that were collected via an operationally viable technique but for which the chemical traces of mines had been verified by analytic chemistry methods. Consequently, we did not fully assess the effectiveness of our generalization training methods. It is likely, however, that the most effective and most efficient strategy for facilitating stimulus generalization will be complex and will involve a blend of the principles described above. Furthermore, we learned that useful assessments of our R&D required further R&D of the sampling phase of REST by chemists and engineers. Once sampling devices and methods are optimized so as to capture maximum chemical evidence of mines and UXO, that technology could be used to ensure that test samples are indeed similar to those that could be obtained in operations. A generalization testing protocol using those test samples could then be developed for repeated use during the course of training.

### CONCLUDING REMARKS

Many readers will be wondering whether this application of ABA was successful. Unfortunately, I cannot affirm definitively that it was because I resigned from my position at GICHD before our R&D program matured to a point at which we could claim either that our system was ready for operational activity or that such a system was not feasible. In addition, the experts who were hired to replace me abandoned a behavior-analytic approach, reverted to the development methods used previously, and subsequently lost funding. Nevertheless, having applied the principles of ABA over just a 2-year period and with only six dogs generated several valuable outcomes. First, we discovered several phenomena and obtained empirical evidence for the utility of numerous training variables, all of which are relevant to training animals in a range of odor-detection roles. (A description of some of our research will soon be disseminated in a publication by GICHD.) Second, my oral presentations of our research likely contributed to GICHD's agreement to sponsor a behavior analyst to work with APOPO on mine detection by giant African pouched rats. That behavior analyst, Alan Poling, has recently published several articles describing APOPO's work in a range of journals (e.g., Poling, Cox, et al., 2010; Poling, Weetjens, Cox, Beyene, Bach, et al., 2010; Poling, Weetjens, Cox, Beyene, & Sully, 2010; Poling, Weetjens, Cox, Beyene, & Sully, 2011; Poling, Weetjens, Cox, Beyene, Durgin, et al., 2011) and has effectively advertised behavior analysis to the humanitarian deminers and humanitarian demining to behavior analysts.

Finally, several authors of recent articles in *The Behavior Analyst* have encouraged readers to pursue novel applications of behavior-analytic research methods so that their utility in the solution of a wide range of

societal concerns might be better recognized (e.g., Critchfield, 2011; Friman, 2010; Poling, 2010). The use of these methods for the development of training procedures in animal-based REST is, I believe, such an application. However, this work represents just a small step toward Skinner's (1953) vision of behavior analysis as a generic science of behavior that is useful for understanding all aspects of the behavior of all living organisms. Perhaps, therefore, teachers of behavior analysis should describe its range of applications while cognizant of the fact that they are describing only a history of published applications and that many more are available for discovery. Furthermore, to position our students to identify, create, and "sell" those yet-to-be-discovered applications, perhaps we should emphasize a distinction between behavior analysis as the generic science of behavior (Skinner, 1953) and behavior-change technologies than can be developed when the principles of this science are applied to benefit some group of people. (See Johnston, 1996, for a discussion of the distinction between applied research and practice.) Put another way, behavior analysis is not solely, and should not be taught solely as, a set of interventions for changing some person's or some animal's behavior, because those interventions are only the technological products of having applied the science of behavior. Instead, the main reason for teaching specific domains of application should be either to prepare someone for using a specific technology (e.g., early intensive behavioral intervention for children with an autism spectrum disorder) or to illustrate, by way of example, the creative process of developing a behavioral technology.

### REFERENCES

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied

- behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20, 313–327.
- Cablk, M. E., & Heaton, J. S. (2006). Accuracy and reliability of dogs in surveying for desert tortoise (*Gopherus agassizii*). *Ecological Applications*, 16(5), 1926–1935.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.
- Critchfield, T. (2011). Translational contributions of the experimental analysis of behavior. *The Behavior Analyst*, 34, 3–17.
- Farthing, G. W., & Hearst, E. (1970). Attention in the pigeon: Testing with compounds and elements. *Learning and Motivation*, 1, 65–78.
- Fjellanger, R. (2003). General principles for training REST dogs. In I. G. McLean (Ed.), *Mine detection dogs: Training, operations, and odor detection* (pp. 11–22). Geneva: Geneva International Center for Humanitarian Demining.
- Fjellanger, R., Andersen, E. K., & McLean, I. G. (2002). A training program for filter-search mine detection dogs. *International Journal of Comparative Psychology*, 15, 278–287.
- Friman, P. C. (2010). Come on in, the water is fine: Achieving mainstream relevance through integration with primary medical care. *The Behavior Analyst*, 33, 19–36.
- Goodwin, K. M., Engel, R. E., & Weaver, D. K. (2010). Trained dogs outperform human surveyors in the detection of rare spotted knapweed (*Centaurea stoebe*). *Invasive Plant Science and Management*, 3(2), 113–121.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Herrnstein, R. J., & Loveland, D. H. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 2, 285–302.
- Hilliard, S. (2003). Principles of animal learning. In I. G. McLean (Ed.), *Mine detection dogs: Training, operations, and odor detection* (pp. 23–42). Geneva: Geneva International Center for Humanitarian Demining.
- Johnson, D. F. (1970). Determiners of selective stimulus control in the pigeon. *Journal of Comparative and Physiological Psychology*, 70, 298–307.
- Johnston, J. M. (1996). Distinguishing between applied research and practice. *The Behavior Analyst*, 19, 35–47.
- Johnston, J. M., Williams, M., Waggoner, L. P., Edge, C. C., Dugan, R. E., & Hallowell, S. F. (1998). Canine detection odor signatures for mine-related explosives. *Proc. SPIE, Detection and Remediation Technologies for Mines and Mine-like Targets III*, 3392, 490–501.
- Joynt, V. (2003). Mechem explosive and drug detection system. In Geneva International Center for Humanitarian Demining (Ed.), *Mine detection dogs: Training, operations, and odor detection* (pp. 165–174). Geneva: Author.
- Kauhanen, E., Harri, M., Nevalainen, A., & Nevalainen, T. (2002). Validity of detection of microbial growth in buildings by trained dogs. *Environment International*, 28, 153–157.
- Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings*. (2nd ed.). New York: Oxford University Press.
- Kiddy, C. A., Mitchell, D. S., Bolt, D. J., & Hawk, H. W. (1978). Detection of estrus-related odors in cows by trained dogs. *Biology of Reproduction*, 19, 389–395.
- Laing, D. G., Panhuber, H., & Slotnick, B. M. (1989). Odor masking in the rat. *Physiology and Behavior*, 45, 689–694.
- Lorenzo, N., Wan, T. L., Harper, R. J., Hsu, Y. L., Chow, M., Rose, S., et al. (2003). Laboratory and field experiments to identify *Canis lupus* var. *familiaris* active odor signature chemicals from drugs, explosives and humans. *Analytical and Bioanalytical Chemistry*, 376, 1212–1224.
- McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *The Behavior Analyst*, 15, 89–94.
- McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst*, 26, 195–213.
- McLean, I. G. (Ed.). (2003). *Mine detection dogs: Training, operations and odour detection*. Geneva: Geneva International Centre for Humanitarian Demining.
- Miles, C. G., & Jenkins, H. M. (1973). Overshadowing in operant conditioning as a function of discriminability. *Learning and Motivation*, 4, 11–27.
- Phelan, J., & Barnett, J. L. (2002). Chemical sensing thresholds for mine detection dogs. In F. L. Orlando, A. C. Dubey, J. F. Harvey, & J. T. Broach (Eds.), *Proceedings of the SPIE 16th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Detection and Remediation Technologies for Mines and Minelike Targets*, 7, 532–543.
- Poling, A. (2010). Looking to the future: Will behavior analysis survive and prosper? *The Behavior Analyst*, 33, 7–17.
- Poling, A., Cox, C., Weetjens, B. J., Beyene, N. W., & Sully, A. (2010). Two strategies for landmine detection by giant pouched rats. *The Journal of ERW and Mine Action*, 14.1.
- Poling, A., Weetjens, B. J., Cox, C., Beyene, N., Bach, H., & Sully, A. (2010). Teaching

- giant African pouched rats to find landmines: Operant conditioning with real consequences. *Behavior Analysis in Practice*, 3, 19–25.
- Poling, A., Weetjens, B., Cox, C., Beyene, N., Durgin, A., & Mahoney, A. (2011). Tuberculosis detection by giant African pouched rats. *The Behavior Analyst*, 34, 47–54.
- Poling, A., Weetjens, B., Cox, C., Beyene, N., & Sully, A. (2010). Using giant African pouched rats (*Cricetomys gambianus*) to detect landmines. *The Psychological Record*, 60, 715–727.
- Poling, A., Weetjens, B. J., Cox, C., Beyene, N. W., Bach, H., & Sully, A. (2011). Using trained pouched rats to detect land mines: Another victory for operant conditioning. *Journal of Applied Behavior Analysis*, 44, 351–355.
- Porter, D., & Neuringer, A. (1984). Musical discriminations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 138–148.
- Pryor, K. (2002). *Don't shoot the dog! The new art of teaching and training*. Gloucestershire, UK: Ringpress Books.
- Quaife, L. R., Moynihan, K. J., & Larson, D. A. (1992). A new pipeline leak-location technique utilizing a novel (patented) test fluid and trained domestic dogs. *Proceedings of the Annual Convention of the Gas Processors Association*, 71, 154–164.
- Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, 4, 203–208.
- Ruszh, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis*, 14, 131–140.
- Schoon, G. A. A. (1998). A first assessment of the reliability of an improved scent identification line-up. *Journal of Forensic Sciences*, 43, 70–75.
- Shelby, R. A., Schrader, K. K., Tucker, A., Klesius, P. H., & Myers, L. J. (2004). Detection of catfish off-flavour compounds by trained dogs. *Aquaculture Research*, 35, 888–892.
- Sidman, M. (1958). Some notes on “bursts” in free operant avoidance experiments. *Journal of Comparative Psychology*, 51, 167–172.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Slotnick, B. M., & Katz, H. M. (1974). Olfactory learning-set formation in rats. *Science*, 185, 796–798.
- Stebbins, W. C. (1970). *Animal psychophysics: The design and conduct of sensory experiments*. New York: Appleton-Century-Crofts.
- Stokes, T. F., & Baer, D. M. (1970). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10, 349–367.
- Waggoner, L. P., Jones, M., Williams, M., Johnston, J. M., Edge, C., & Petrousky, J. A. (1998). Effects of extraneous odors on canine detection. *Proceedings of the 2nd annual conference on Enforcement and Security Technologies*. Bellingham, WA: The International Society for Optical Engineering.
- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63, 165–174.
- Williams, H., & Pembroke, A. (1989). Sniffer dogs in a melanoma clinic? *Lancet*, 1, 734.
- Williams, M., & Johnston, J. M. (2002). Training and maintaining the performance of dogs (*Canis familiaris*) on an increasing number of odor discriminations in a controlled setting. *Applied Animal Behavior Science*, 78, 55–65.